International Psychogeriatrics: page 1 of 8 © International Psychogeriatric Association 2017
doi:10.1017/S1041610217001417

Measurement invariance across educational levels and gender in 12-item Zarit Burden Interview (ZBI) on caregivers of people with dementia

Chung-Ying Lin,[1] Li-Jung Elizabeth Ku[2] and Amir H. Pakpour[3]

[1]*Department of Rehabilitation Sciences, Faculty of Health and Social Sciences, Hong Kong Polytechnic University, Hung Hom, Hong Kong*
[2]*Department of Public Health, National Cheng Kung University Hospital, College of Medicine, National Cheng Kung University, Tainan, Taiwan*
[3]*Social Determinants of Health Research Center, Qazvin University of Medical Sciences, Qazvin, Iran; Department of Nursing, School of Health and Welfare, Jönköping University, Jönköping, Sweden*

ABSTRACT

Background: The Zarit Burden Interview (ZBI) is a commonly used self-report to assess caregiver burden. A 12-item short form of the ZBI has been developed; however, its measurement invariance has not been examined across some different demographics. It is unclear whether different genders and educational levels of a population interpret the ZBI items similarly. Therefore, this study aimed to examine the measurement invariance of the 12-item ZBI across gender and educational levels in a Taiwanese sample.

Methods: Caregivers who had a family member with dementia (n = 270) completed the ZBI through telephone interviews. Three confirmatory factor analysis (CFA) models were conducted: Model 1 was the configural model, Model 2 constrained all factor loadings, Model 3 constrained all factor loadings and item intercepts. Multiple group CFAs and the differential item functioning (DIF) contrast under Rasch analyses were used to detect measurement invariance across males (n = 100) and females (n = 170) and across educational levels of junior high schools and below (n = 86) and senior high schools and above (n = 183).

Results: The fit index differences between models supported the measurement invariance across gender and across educational levels (Δ comparative fit index (CFI) = −0.010 and 0.003; Δ root mean square error of approximation (RMSEA) = −0.006 to 0.004). No substantial DIF contrast was found across gender and educational levels (value = −0.36 to 0.29).

Conclusions: The ZBI is appropriate for combined use and for comparisons in caregivers across gender and different educational levels in Taiwan.

Key words: burden, caregiving, dementia, gender, education, measurement invariance

Introduction

Given the extended life expectancy worldwide, the elderly people have more possibilities to encounter different health problems, especially chronic diseases. One of the most serious chronic disease disorders for geriatrics is dementia (Douzenis, 2008). Indeed, people with dementia (PWD) whose cognitive abilities gradually decline often face declining functional performance during the disease progression (Marcotte and Grant, 2010). Furthermore, impaired cognitive ability and declining functional performance are substantial challenges for individuals who take care of their beloved family member who has dementia. Such caregivers need to provide continuous day-to-day supervision and care, and to tolerate and cope with the inappropriate behaviors from their family member with dementia (Ankri *et al.*, 2005). Therefore, caregivers may not be able to collaborate well with healthcare providers (e.g. neurologists, psychiatrists, nurses, social workers, and occupational therapists) to treat their family members with dementia, and may even jeopardize their own health (Schreiner *et al.*, 2006; Lin *et al.*, 2017). The caregivers may suffer burnout and have no energy to take care of their family members with dementia. To protect caregivers from burnout and the risk of becoming ill, healthcare providers need to assess the caregiver burden using a

Correspondence should be addressed to: C.-Y. Lin, PhD, Department of Rehabilitation Sciences, Faculty of Health and Social Sciences, Hong Kong Polytechnic University, 11 Yuk Choi Rd, Hung Hom, Kowloon, Hong Kong. Phone: +852-2766-6755; Fax: +852-2764-1435. Email: cylin36933@gmail.com. Received 23 Nov 2016; revision requested 13 Jan 2017; revised version received 13 Feb 2017; accepted 28 Jun 2017.

concise and short patient-reported outcome (PRO) measure.

A PRO measure helps healthcare providers understand the health status from the perspective of each individual, and is extremely important when measuring those health outcomes that cannot be assessed through objective techniques, such as burden and pain. Moreover, in 2009 the US Food and Drug Administration (FDA) recommended the usefulness of PRO measures so long as the measure has sound psychometric properties. Some researchers further suggest that a well-established PRO measure that assesses the feelings of an individual can greatly help healthcare providers make appropriate clinical decisions (Chang *et al.*, 2014). Thus, we recommend that healthcare providers use Zarit Burden Interview (ZBI) – a commonly used PRO measure to assess caregiver burden (Branger *et al.*, 2014) – to evaluate the burden of caregivers of family members with dementia.

However, when using the ZBI to capture caregiver burden, a critical and practical question for healthcare providers is whether respondents of different genders or different educational levels interpret the PRO items differently. That is, we should identify whether the score difference of a measure reflects the true group differences or is due to various understandings toward the same measure contents (Gregorich, 2006). For example, some studies have found that female caregivers experience higher levels of stigma and burden than male caregivers (Hayden and Heller, 1997; Lai, 2012; Kahn *et al.*, 2016). One study also indicated that family caregivers with a higher educational level may have a higher level of caregiving burden (Lai, 2012). With such information, healthcare providers may foster effective intervention based on the special needs for different genders or educational levels. Thus, healthcare providers and researchers may want to compare the differences between subgroups (e.g. males vs. females) to help them make the best clinical decisions (Lin *et al.*, 2016). However, without examining the measurement invariance, we cannot ensure the differences are real differences or determine whether the differences are due to different interpretations. Hence, considering the measurement invariance across the subgroups is a must from the view of psychometric testing.

Unfortunately, except for one study on the measurement invariance across Black and White caregivers (Longmire and Knight, 2011), no studies have explored the measurement invariance issues for the ZBI. Therefore, given that differences of burden levels were found even between genders (Hayden and Heller, 1997; Lai, 2012; Kahn *et al.*, 2016) or between educational levels (Lai, 2012), healthcare providers cannot ensure whether the differences are true or only reflect different interpretations of the item contents. Taking the ZBI item *not enough time* as an example, the difference in the item score could be because different genders (or different educational levels) really devote their caring time differently. However, we cannot rule out another possibility: that men and women (or different educational levels) may have different perceptions of "time." If men interpret 40 h a week as not enough and females 20 h a week as not enough, comparing the item scores between genders would be inappropriate. Therefore, before using the ZBI scores for comparisons, we need to investigate whether genders (or different educational levels) interpret ZBI items similarly.

Two statistical methods can be applied to examine the measurement invariance: confirmatory factor analysis (CFA) or Rasch analysis. Measurement invariance is the term often used in CFA and is used less often in Rasch analysis. In Rasch analysis, the commonly used term is differential item functioning (DIF) that shares the same meaning with measurement invariance in CFA. Thus, some researchers exchangeably use measurement invariance and DIF (Jafari *et al.*, 2014) whether using CFA or Rasch analysis. Because we attempted to use the two methods to detect the measurement invariance of the ZBI items, we used the term DIF solely for Rasch analysis and the term *measurement invariance majority* for CFA. That is, we also used the term *measurement invariance* to indicate the broad concept of the equivalent interpretations between subgroups, and did not specifically indicate which kinds of statistical methods were used. Using different methods to test the same topic helps us to determine whether the evidence on measurement invariance is robust. If both methods support the measurement invariance, healthcare providers can use the ZBI without hesitation; conversely, if the results are diverse between methods, healthcare providers may use the ZBI with caution.

The purpose of this study was to use both CFA and Rasch analysis to investigate measurement invariance (or say, DIF) across gender and educational levels for the 12-item ZBI in a sample of Taiwanese caregivers. We also separated the educational level into $\leq$ junior high versus $\geq$ senior high because compulsory education in Taiwan is up to junior high. In addition, the years of education were 9 for junior high and 12 for senior high. The educational level of participants who were enrolled in but who had not completed studying in senior high, was defined as junior high.

Methods

Participants and procedure

Between November 2013 and April 2015, 286 dementia patient-caregiver dyads were recruited from the dementia clinic in a national university hospital in southern Taiwan. Ethical approval was obtained from the National Cheng Kung University Hospital Institutional Review Board before this study began (IRB No: B-ER-102-173). The inclusion criteria for PWD were as follows: (1) with a diagnosis of dementia according to the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV), (2) living in the community, and (3) having an unpaid caregiver. Dementia subtypes included mostly Alzheimer's disease ($n = 224$), but also dementia with Lewy bodies ($n = 9$), vascular dementia ($n = 6$), and other unspecified types. Inclusion criteria for the caregiver of PWD were as follows: (1) he or she must be a family member who had cared for the PWD for at least 12 months, (2) be $\geq$18 years of age, and (3) fluent in either Mandarin or Taiwanese. After obtaining written consent for the study from the caregivers and the PWD, phone interviews of the caregivers were conducted. For cognitively impaired PWD who could not provide their signatures, proxy consent was obtained from their family caregiver. After excluding 10 PWD who moved to a nursing home, 4 withdrawals, and 2 incomplete responses, our final sample included 270 community-dwelling dyads.

Using the questionnaires, we collected data on baseline demographics (age, gender, and educational level), marital, and employment status from caregivers. The relationship to the care-recipient was categorized into spouse, adult child, adult child-in-law, and other. Unpaid caregiver time provided by the family caregiver was also collected. Care hours were categorized into hours spent on Activities of Daily Living (ADL), Instrumental Activities of Daily Living (IADL), and supervision of the person with dementia in the past week before the interview. Care duration measured how many years the caregiver had been taking care of the care-recipient.

Zarit Burden Interview (ZBI)

There were several versions of the ZBI, including 1- to 5-factor structures with items varying from 12 to 22 (Ballesteros *et al.*, 2012; Cheng *et al.*, 2014; Lau *et al.*, 2015). The original version containing 22 items has been translated into a Chinese version for Taiwanese with excellent test–retest reliability (intraclass correlation coefficient = 0.88) and internal consistency ($\alpha = 0.89$) (Ko *et al.*, 2008). The Chinese version used in Taiwan has 22 items spread in five factors: feelings of over-sacrifice, PWD's dependence, negative emotion due to caregiving, feeling of inadequacy, and uncertainty about PWD's future.

Instead of using the 22-item version, we examined measurement invariance using a short version with 12 items embedded in an overall latent construct (i.e. burden) because this short version has been found to outperform other versions (Ballesteros *et al.*, 2012; Lin *et al.*, 2017). Its internal consistency and concurrent validity have also been found to be satisfactory (Lin *et al.*, 2017).

Data analysis

Measurement invariance and CFA models were performed using LISREL 8.8 (Scientific Software International, Lincolnwood, IL), DIF, and Rasch models using WINSTEPS 3.75.0, and all other analyses using SPSS 16.0 (SPSS Inc., Chicago, IL).

Three models were used for measurement invariance. Model 1 was the configural model (i.e. a first-order ZBI framework with all 12 items loaded on the underlying burden concept); Model 2 constrained all factor loadings; Model 3 additionally constrained all item intercepts. In addition, the three models were applied to gender for gender invariance and to educational level for invariance across educational level. For the configural model, non-significant χ^2, comparative fit index (CFI) > 0.95, standardized root mean square residual (SRMR) < 0.08, and root mean square error of approximation (RMSEA) < 0.08 support the first-order ZBI framework (Jafari *et al.*, 2012b). Measurement invariance was assessed by comparing the nested Models 1 to 3 (Bagheri *et al.*, 2014), that is, $\Delta\chi^2$, ΔCFI, ΔSRMR, and ΔRMSEA. Non-significant $\Delta\chi^2$, $\Delta\text{CFI} > -0.01$, $\Delta\text{SRMR} < 0.01$, and $\Delta\text{RMSEA} < 0.015$ (Chen, 2007) suggest measurement invariance, while χ^2 and $\Delta\chi^2$ have the shortcoming of being too easily to be statistically significant in a large sample size.

Before assessing DIF, we used a Rasch partial credit model (PCM) to retest the unidimensionality of the entire ZBI and the item fit. The item fit was assessed using both the information-weighted fit statistic (infit) mean square (MNSQ) and the outlier-sensitive fit statistic (outfit) MNSQ. Both ranges between 0.5 and 1.5 suggest well fit (Jafari *et al.*, 2012a): a very low MNSQ suggests a redundant item and a very high MNSQ suggests an out-of-concept item (Khan *et al.*, 2013). DIF analysis was conducted to corroborate the measurement invariance assessed using CFA, and a DIF item suggests that it is too difficult or too easy for a particular group (different gender

Table 1. Demographics and Zarit Burden Interview score of caregivers ($N = 270$)

	GENDER		EDUCATIONAL LEVEL	
	MALE (N = 100)	FEMALE (N = 170)	JUNIOR HIGH AND BELOW[a] (N = 86)	SENIOR HIGH AND ABOVE[b] (N = 183)
Age (years), M (SD)	63.66 (14.37)	57.82 (12.02)	68.80 (11.37)	55.94 (11.96)
Marital status (married), *n* (%)	88 (88.0)	132 (77.6)	82 (95.3)	137 (74.9)
Employment (yes), *n* (%)	48 (48.0)	70 (41.2)	26 (30.2)	92 (50.3)
Relationship to care-recipient, *n* (%)				
Spouse	46 (46.0)	63 (37.1)	69 (80.2)	40 (21.9)
Adult child	52 (52.0)	72 (42.4)	12 (14.0)	111 (60.7)
Adult child-in-law	1 (1.0)	31 (18.2)	5 (5.8)	27 (14.8)
Others	1 (1.0)	4 (2.4)	0 (0.0)	5 (2.7)
Living with care-recipient, *n* (%)	83 (83.0)	127 (74.7)	82 (95.3)	128 (69.9)
Care hours per week, M/SD				
Care for activity of daily living (ADL)	6.99 (9.13)	6.34 (8.84)	7.51 (9.61)	6.15 (8.61)
Care for instrumental ADL (IADL)	12.20 (9.98)	16.51 (9.53)	17.95 (9.56)	13.49 (9.79)
Care for ADL and IADL	19.20 (16.09)	22.84 (15.38)	25.47 (15.25)	19.63 (15.66)
Care for ADL, IADL, and supervision	47.05 (42.06)	49.87 (40.83)	61.30 (41.35)	43.01 (40.08)
Care duration (years), M (SD)[c]	5.36 (3.31)	4.21 (2.88)	4.75 (3.07)	4.57 (3.12)
Zarit Burden Interview[d]	1.37 (0.81)	1.47 (0.88)	1.36 (0.90)	1.47 (0.84)

[a]Educational year of junior high in Taiwan is 9, and the educational level of those who have enrolled but not completed studying in senior high were defined as junior high.
[b]Educational year of senior high in Taiwan is 12.
[c]With one missing value.
[d]Reported using four-point Likert scale.

and different educational level in this study) of respondents (Lin *et al.*, in press). We used DIF contrasts (i.e. the difference of difficulty between two groups); a value > 0.5 suggests a substantial DIF (Shih and Wang, 2009).

Results

There were 100 males and 170 females in this study, and the majority of the caregivers were primary caregivers ($n = 232$; 85.9%). One of the caregivers did not report her educational level, 86 had an educational level of junior high and below; 183 had an educational level of senior high and above. The demographics separated by gender and educational level are presented in Table 1. Specifically, male caregivers were older than females (Difference±SD = 5.84 ± 1.71, p = 0.001), and had fewer care hours for IADL (Difference±SD = −4.30 ± 1.2, p = 0.001) and longer care duration (Difference±SD = 1.15 ± 0.40, p = 0.005). Caregivers with an educational level of junior high and below were much older than those of senior high and above (Difference ± SD = 12.86 ± 1.54, p <0.001), and had more care hours for IADL (Difference±SD = 4.47 ± 1.27, p = 0.001) and total care (Difference±SD = 18.28 ± 5.29, p = 0.001). Most caregivers with an educational level of junior high and below were spouses of their care-recipient (80.2%), while those of senior high and above were children of the care-recipient (60.7%).

Two configural models (one for gender and the other for educational level) had satisfactory fit indices (CFI = 0.959 and 0.955, SRMR = 0.052 and 0.063, and RMSEA = 0.077 and 0.078), except for the significant χ^2. After constraining all factor loadings as equal across gender and educational level, all the models still had excellent fit statistics (CFI = 0.962 and 0.955, SRMR = 0.057 and 0.068, and RMSEA = 0.071 and 0.073), except for the significant χ^2. Identical results were found when we additionally constrained all item intercepts being equal across gender and educational levels (CFI = 0.961 and 0.945, SRMR = 0.057 and 0.066, RMSEA = 0.068 and 0.077, and significant χ^2). The model comparisons suggest that measurement invariance was supported across gender and across educational levels based on non-significant χ^2 difference tests (only one significant χ^2 difference test was shown in our results) and other fit indices (ΔCFI = −0.001 to 0.003, ΔSRMR = −0.002 to 0.005, and ΔRMSEA = −0.006 to 0.004) (Table 2).

Results of Rasch analyses are presented in Table 3. In brief, the easiest item for caregivers was

Table 2. Measurement invariance across educational level and across gender

MODEL AND COMPARISONS	FIT STATISTICS							
	χ^2 (df)	$\Delta\chi^2$ (Δdf)	CFI	ΔCFI	SRMR	ΔSRMR	RMSEA	ΔRMSEA
Gender								
M1: Configural	190.602 (108)*		0.959		0.052		0.077	
M2: Plus all loadings constrained	196.291 (119)*		0.962		0.057		0.071	
M3: Plus all intercepts constrained	209.623 (130)*		0.961		0.057		0.068	
M2−M1		5.689 (11)		0.003		0.005		−0.006
M3−M2		13.332 (11)		−0.001		0.000		−0.003
Educational level								
M1: Configural	198.370 (108)*		0.955		0.063		0.078	
M2: Plus all loadings constrained	209.088 (119)*		0.955		0.068		0.073	
M3: Plus all intercepts constrained	240.746 (130)*		0.945		0.066		0.077	
M2−M1		10.718 (11)		0.000		0.005		−0.005
M3−M2		31.658 (11)*		−0.010		−0.002		0.004

*$p < 0.05$.
CFI = comparative fit index; SRMR = standardized root mean square residual; RMSEA = root mean square error of approximation.

Table 3. Differential item functioning (DIF) for 12-Item Zarit Burden Interview (ZBI)

Item #	Difficulty	MnSq		DIF contrast	
		Infit	Outfit	Gender[a]	Education[b]
2. Not enough time	0.24	0.96	0.86	0.14	0.10
3. Stress	−0.13	0.70	0.68	0.11	−0.02
8. Relative's dependence	−0.57	1.55	1.85	0.00	0.25
9. Strained by relative	0.76	1.13	1.21	−0.15	−0.19
10. Health decrease	0.14	0.89	0.98	−0.22	0.12
11. Lack of privacy	−0.03	0.81	0.82	−0.36	−0.18
12. Lack of social life	−0.06	0.95	0.95	−0.10	−0.09
16. Unable to care much longer	−0.39	1.52	1.86	0.06	0.14
17. Lost control of life	−0.11	0.77	0.68	0.29	0.00
18. Leave the care to others	0.48	1.05	1.33	−0.08	−0.12
19. Uncertain about what to do	0.40	1.06	1.06	0.29	−0.18
22. Overall burden	−0.74	0.68	0.66	0.00	0.00

MnSq = Mean Square; Infit = information-weighted fit statistic; Outfit = outlier-sensitive fit statistic.
[a]The difference of difficulty on male–difficulty on female.
[b]The difference of difficulty on caregivers educated junior high and below–difficulty on caregivers educated senior high and above.

the overall burden (Item 22, difficulty = −0.74), and the hardest item was strained by relatives (Item 9, difficulty = 0.76). In addition, all items had acceptable or nearly acceptable Infit MNSQ values, but 2 items (Item 8: relative's dependence, and Item 16: unable to care much longer) had unsatisfactory Outfit MNSQ (1.85 and 1.86, respectively). Neither item displayed a substantial DIF across gender (DIF contrast = −0.36 to 0.29) nor educational level (DIF contrast = −0.19 to 0.25).

Discussion

Because the ZBI is a commonly used measure for assessing caregiver burden, many studies regarding caregivers of PWD examine the psychometric properties of the ZBI in the specific group (Ballesteros *et al.*, 2012; Cheng *et al.*, 2014; Lau *et al.*, 2015). These studies provide constructive and useful information for healthcare providers in terms of assessing caregiver burden for the suffering population, and our results extend the

knowledge to the measurement invariance of the ZBI.

The primary benefit of using the 12-item ZBI is its efficiency (Bédard *et al.*, 2001); it only takes less than 10 min for a caregiver to fill out. In addition, the precision of the 12-item ZBI is aligned with the original 22-item ZBI, as indicated by one study that showed a relatively high correlation ($r = 0.952$) between the 12-item ZBI score and the 22-item ZBI score (Lin *et al.*, 2017). The unidimensional structure of the 12-item ZBI has been supported by both CFA and Rasch analysis; the internal consistency of the 12-item ZBI is excellent (Ballesteros *et al.*, 2012; Lin *et al.*, 2017). Also, the 12-item ZBI score can efficiently differentiate caregivers who are in a good financial situation and those in a poor financial situation (Lin *et al.*, 2017). Hence, we are confident that the 12-item ZBI has the ability to help healthcare providers to well understand the caregiver burden, and to quickly foster appropriate intervention plans, such as providing respite care (Greenwood *et al.*, 2012).

Based on the sound psychometric properties of the 12-item ZBI, we elaborated on the measurement invariance issue of the 12-item ZBI. Although one study found that the ZBI was invariant across Black and White dementia caregivers (Longmire and Knight, 2011), to the best of our knowledge, no studies have examined the issue across gender and educational level. Our results showed that all fit indices, including χ^2 difference tests, ΔCFI, ΔSRMR, and ΔRMSEA, support the measurement invariance across male and female caregivers. Except for a significant χ^2 difference test, all the fit indices testing for measurement invariance suggest that caregivers in junior high educational level or lower view the ZBI item contents the same as those at senior high educational level or higher. Because the χ^2 difference test is too sensitive when the sample size is large and might not be reliable (Chen *et al.*, 2005), we hereby claim that our results fully support the measurement invariance of the 12-item ZBI. In addition, our results of Rasch analysis displayed no substantial DIF across gender and educational levels, which corroborated our CFA results. Therefore, we consider that we have sufficient and robust results to claim the measurement invariance for the 12-item ZBI. With the support of measurement invariance, healthcare providers and researchers may use the 12-item ZBI to compare the caregiver burden between genders or between different educational levels.

There are several limitations in this study. First, all our participants were recruited from the same dementia clinic and our results may not be generalizable to the entire population of Taiwan. Second, because our study sample included only caregivers of community-dwelling dementia patients, our results may not be generalizable to dementia patients living in institutions. Given that dementia patients in advanced stages are more likely to be institutionalized, our sample mainly reflects caregiver burden for patients in mild or moderate stages of dementia. Third, the cultural issue: healthcare providers and researchers should be informed that the 12-item ZBI does not address a salient concept in Asian societies – the worry about performance. This limitation should be paid attention to because filial piety is an esteemed value in East Asia (Lim *et al.*, 2014). Hence, our findings may not be generalizable to Western countries, and future studies in the West are warranted. Fourth, we did not have information regarding the caregivers' depression or anxiety disorders. Caregivers of PWD often suffer from depression or anxiety, and such information may shed some light on the characteristics of our cohort of caregivers. Therefore, we suggest that future studies on the same topic may want to measure the caregivers' depression and anxiety. Finally, our sample size of caregivers with an educational level of junior high and below was not large ($n = 86$). Therefore, the findings of measurement invariance across educational levels should be interpreted with caution. That is, future studies with a large sample are warranted to corroborate our findings.

Our results provide important scientific and clinical implications because measuring caregiver burden is a crucial issue for PWD. From a scientific point of view, an assessment of caregiver burden with supported measurement invariance may help researchers compare the burden levels among different groups. Further information and knowledge may thus be established for the population. Our results also provided the rationale and justification for studies that have used ZBI to compare caregiver burden between genders or educational levels.

From a clinical point of view, when a healthcare provider wants to know the differences of burden between male and female caregivers (or caregivers with higher and lower educational levels) of the same patient, our results confirm that using the ZBI is appropriate. That is, our results assure healthcare providers that it is meaningful to compare the caregiver burden across gender or educational levels in such a population. The comparisons are supported because our results indicate that different genders and caregivers in different educational levels interpret the ZBI item content in the same way. Because the measurement invariance was supported, healthcare providers may comfortably use the results of other

studies for comparing ZBI between genders and educational levels to design interventions. This is crucial because gender and educational levels are important factors in caregiver burden (Hayden and Heller, 1997; Lai, 2012; Kahn *et al.*, 2016), and caregiver burden may subsequently impact the coping skills and health of the caregivers (Parekh *et al.*, 2016).

Conclusions

The 12-item ZBI is a promising PRO measure to assess the caregiver burden for caregivers taking care of family members with dementia. In addition to its efficiency and sound psychometric properties, we found that the 12-item ZBI can be used to compare the caregiver burden between male and female caregivers, and between caregivers with an educational level of junior high or lower and those with an educational level of senior high or above. Healthcare providers may want to use the short and concise measure to properly evaluate the caregiver burden and use the information it measured to design appropriate interventions.

Conflict of interest

None.

Description of authors' roles

CY Lin and LJE Ku designed the research questions. CY Lin wrote the first draft and analyzed the data. LJE Ku collected the data and critically revised the manuscript. AH Pakpour assisted in the data analyses, helped interpret the statistical results, and critically revised the manuscript.

Acknowledgments

This research was supported by grants from the National Science Council (NSC 102-2410-H-006-001-SS2). The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

Ankri, J., Andrieu, S., Beaufils, B., Grand, A. and Henrard, J. C. (2005). Beyond the global score of the Zarit Burden Interview: useful dimensions for clinicians. *International Journal of Geriatric Psychiatry*, 20, 254–260. doi:10.1002/gps.1275.

Bagheri, Z., Jafari, P., Tashakor, E., Kouhpayeh, A. and Riazi, H. (2014). Assessing whether measurement invariance of the KIDSCREEN-27 across child-parent dyad depends on the child gender: a multiple group confirmatory factor analysis. *Global Journal of Health Science*, 6, 142–153. doi:10.5539/gjhs.v6n5p142.

Ballesteros, J., Santos, B., González-Fraile, E., Muñoz-Hermoso, P., Domínguez-Panchón, A. I. and Martín-Carrasco, M. (2012). Unidimensional 12-item Zarit Caregiver Burden Interview for the assessment of dementia caregivers' burden obtained by item response theory. *Value in Health*, 15, 1141–1147. doi:10.1016/j.jval.2012.07.005.

Bédard, M., Molloy, D. W., Squire, L., Dubois, S., Lever, J. A. and O'Donnell, M. (2001). The Zarit Burden Interview: a new short version and screening version. *Gerontologist*, 41, 652–657. doi:10.1093/geront/41.5.652.

Branger, C., O'Connell, M. E. and Morgan, D. G. (2014). Factor analysis of the 12-item Zarit Burden Interview in caregivers of persons diagnosed with dementia. *Journal of Applied Gerontology*, 35, 489–507. doi:10.1177/0733464813520222.

Chang, K.-C., Wang, J.-D., Tang, H.-P., Cheng, C.-M. and Lin, C.-Y. (2014). Psychometric evaluation, using Rasch analysis, of the WHOQOL-BREF in heroin-dependent people undergoing methadone maintenance treatment: further item validation. *Health and Quality of Life Outcomes*, 12, 148. doi:10.1186/s12955-014-0148-6.

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling*, 14, 464–504. doi:10.1080/10705510701301834.

Chen, F. F., Sousa, K. H. and West, S. G. (2005). Teacher's Corner: testing measurement invariance of second-order factor models. *Structural Equation Modeling*, 12, 471–492. doi:10.1207/s15328007sem1203_7.

Cheng, S.-T., Kwok, T. and Lam, L. C. W. (2014). Dimensionality of burden in Alzheimer caregivers: confirmatory factor analysis and correlates of the Zarit Burden interview. *International Psychogeriatrics*, 26, 1455–1463. doi:10.1017/S104161021400101X.

Douzenis, A. (2008). Dementia in the general hospital. *Annals of General Psychiatry*, 7, S21. doi:10.1186/1744-859X-7-S1-S21.

Greenwood, N., Habibi, R. and Mackenzie, A. (2012). Respite: carers' experiences and perceptions of respite at home. *BMC Geriatrics*, 12, 42. doi:10.1186/1471-2318-12-42.

Gregorich, S. E. (2006). Do self-report instruments allow meaningful comparisons across diverse population groups? Testing measurement invariance using the confirmatory factor analysis framework. *Medical Care*, 44, S78–S94. doi:10.1097/01.mlr.0000245454.12228.8f.

Hayden, M. F. and Heller, T. (1997). Support, problem-solving/coping ability, and personal burden of younger and older caregivers of adults with mental retardation. *Mental Retardation*, 35, 364–372. doi:10.1352/0047-6765(1997)035<0364:SPAAPB>2.0.CO;2.

Jafari, P., Bagheri, Z. and Safe, M. (2012a). Item and response-category functioning of the Persian version of the

KIDSCREEN-27: Rasch partial credit model. *Health and Quality of Life Outcomes*, 10, 127. doi:10.1186/1477-7525-10-127.

Jafari, P., Bagheri, Z., Ayatollahi, S. M. T. and Soltani, Z. (2012b). Using Rasch rating scale model to reassess the psychometric properties of the Persian version of the PedsQL^{TM} 4.0 Generic Core Scales in school children. *Health and Quality of Life Outcomes*, 10, 27. doi:10.1186/1477-7525-10-27.

Jafari, P., Sharafi, Z., Bagheri, Z. and Shalileh, S. (2014). Measurement equivalence of the KINDL questionnaire across child self-reports and parent proxy-reports: a comparison between item response theory and ordinal logistic regression. *Child Psychiatry & Human Development*, 45, 369–376. doi:10.1007/s10578-013-0407-5.

Kahn, P. V., Wishart, H. A., Randolph, J. S. and Santulli, R. B. (2016). Caregiver stigma and burden in memory disorders: an evaluation of the effects of caregiver type and gender. *Current Gerontology and Geriatrics Research*, 2016, Article ID 8316045. doi:10.1155/2016/8316045.

Khan, A., Chien, C.-W. and Brauer, S. G. (2013). Rasch-based scoring offered more precision in differentiating patient groups in measuring upper limb function. *Journal of Clinical Epidemiology*, 66, 681–687. doi:10.1016/j.jclinepi.2012.12.014.

Ko, K.-T., Yip, P.-K., Liu, S.-I. and Huang, C.-R. (2008). Chinese version of the Zarit Caregiver Burden Interview: a validation study. *American Journal of Geriatric Psychiatry*, 16, 513–518. doi:10.1097/JGP.0b013e318167ae5b.

Lai, D. W. L. (2012). Effect of financial costs on caregiving burden of family caregivers of older adults. *Sage Open*, 1–14. doi:10.1177/2158244012470467.

Lau, S., Chong, M. S., Ali, N. Chan, M., Chua, K. C. and Lim, W. S. (2015). Caregiver burden: looking beyond the unidimensional total score. *Alzheimer Disease and Associated Disorders*, 29, 338–346. doi:10.1097/WAD.0000000000000085.

Lim, W. S. *et al*. (2014). Worry about performance: a unique dimension of caregiver burden. *International Psychogeriatrics*, 26, 677–686. doi:10.1017/S1041610213002445.

Lin, C.-Y., Li, Y.-P., Lin, S.-I. and Chen, C.-H. (2016). Measurement equivalence across gender and education in the WHOQOL-BREF for community-dwelling elderly Taiwanese. *International Psychogeriatrics*, 28, 1375–1382. doi:10.1017/S1041610216000594.

Lin, C.-Y., Wang, J.-D., Pai, M.-C. and Ku, L.-J. E. (2017). Measuring burden in dementia caregivers: confirmatory factor analysis for short forms of the Zarit Burden Interview. *Archives of Gerontology and Geriatrics*, 68, 8–13. doi:10.1016/j.archger.2016.08.005.

Lin, C.-Y., Yang, S.-C., Lai, W.-W., Su, W.-C. and Wang, J.-D. (2017). Rasch models suggested the satisfactory psychometric properties of the WHOQOL-BREF among lung cancer patients. *Journal of Health Psychology*, 22, 397–408. doi:10.1177/1359105315603474.

Longmire, C. V. and Knight, B. G. (2011). Confirmatory factor analysis of a brief version of the Zarit Burden Interview in Black and White dementia caregivers. *Gerontologist*, 51, 453–462.

Marcotte, T. D. and Grant, I. (2010). *Neuropsychology of Everyday Functioning*. New York: Guilford Press.

Parekh, N. K. *et al*. (2016). Effects of caregiver burden on quality of life and coping strategies utilized by caregivers of adult patients with inflammatory bowel. *Annals of Gastroenterology*, 29, 1–7.

Schreiner, A. S., Morimoto, T., Arai, Y. and Zarit, S. (2006). Assessing family caregiver's mental health using a statistically derived cut-off score for the Zarit Burden Interview. *Aging and Mental Health*, 10, 107–111. doi:10.1080/13607860500312142.

Shih, C.-L. and Wang, W.-C. (2009). Differential item functioning detection using the multiple indicators, multiple causes method with a pure short anchor. *Applied Psychological Measurement*, 33, 184–199. doi:10.1177/0146621608321758.